1. In the sense of machine learning, what is a model? What is the best way to train a Model?

Ans:. When to use Machine Learning

Good machine learning scenarios often have the following common properties:

They involve a repeated decision or evaluation which you want to automate and need consistent results.

It is difficult or impossible to explicitly describe the solution or criteria behind a decision.

You have labeled data, or existing examples where you can describe the situation and map it to the correct result.

Windows Machine Learning uses the Open Neural Network Exchange (ONNX) format for its models. You can download a pre-trained model, or you can train your own model. See Get ONNX models for Windows ML for more information

A learning model is a description of the mental and physical mechanisms that are involved in the acquisition of new skills and knowledge and how to engage those those mechanisms to encourage and facilitate learning. In AI/ML, a model replicates a decision process to enable automation and understanding. AI/ML models are mathematical algorithms that are “trained” using data and human expert input to replicate a decision an expert would make when provided that same information.

2. In the sense of machine learning, explain the “No Free Lunch’ theorem.?

Ans:.

The field of optimization and machine learning, often with little understanding of what it means or implies.

The theorem states that all optimization algorithms perform equally well when their performance is averaged across all possible problems.

It implies that there is no single best optimization algorithm. Because of the close relationship between optimization, search, and machine learning, it also implies that there is no single best machine learning algorithm for predictive modeling problems such as classification and regression.

In this tutorial, you will discover the no free lunch theorem for optimization and search.

After completing this tutorial, you will know:

The no free lunch theorem suggests the performance of all optimization algorithms are identical, under some specific constraints.

There is provably no single best optimization algorithm or machine learning algorithm.

The practical implications of the theorem may be limited given we are interested in a small subset of all possible objective functions.

Kick-start your project with my new book Optimization for Machine Learning, including step-by-step tutorials and the Python source code .

3. Describe the K-fold cross-validation mechanism in detail.?

Ans:. k-Fold Cross-Validation

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample.

The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation.

Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample in order to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model.

It is a popular method because it is simple to understand and because it generally results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.

The general procedure is as follows:

Shuffle the dataset randomly.

Split the dataset into k groups

For each unique group:

Take the group as a hold out or test data set

Take the remaining groups as a training data set

Fit a model on the training set and evaluate it on the test set

Retain the evaluation score and discard the model

Summarize the skill of the model using the sample of model evaluation scores

Importantly, each observation in the data sample is assigned to an individual group and stays in that group for the duration of the procedure. This means that each sample is given the opportunity to be used in the hold out set 1 time and used to train the model k-1 times.

4. Describe the bootstrap sampling method. What is the aim of it?

Ans:. Bootstrap Method

The bootstrap method is a statistical technique for estimating quantities about a population by averaging estimates from multiple small data samples.

Importantly, samples are constructed by drawing observations from a large data sample one at a time and returning them to the data sample after they have been chosen. This allows a given observation to be included in a given small sample more than once. This approach to sampling is called sampling with replacement.

The process for building one sample can be summarized as follows:

Choose the size of the sample.

While the size of the sample is less than the chosen size

Randomly select an observation from the dataset

Add it to the sample

The bootstrap method can be used to estimate a quantity of a population. This is done by repeatedly taking small samples, calculating the statistic, and taking the average of the calculated statistics. We can summarize this procedure as follows:

Choose a number of bootstrap samples to perform

Choose a sample size

For each bootstrap sample

Draw a sample with replacement with the chosen size

Calculate the statistic on the sample

Calculate the mean of the calculated sample statistics.

The procedure can also be used to estimate the skill of a machine learning model.This is done by training the model on the sample and evaluating the skill of the model on those samples not included in the sample. These samples not included in a given sample are called the out-of-bag samples, or OOB for short.

5. What is the significance of calculating the Kappa value for a classification model? Demonstrate how to measure the Kappa value of a classification model using a sample collection of results.?

Ans:.

Cohen’s Kappa is a statistical measure that is used to measure the reliability of two raters who are rating the same quantity and identifies how frequently the raters are in agreement.

In this article, we will learn in detail about what Cohen’s kappa is and how it can be useful in machine learning problems.

Intra-rater and inter-rater reliability

Before we understand Cohen’s Kappa, let us understand what Intra and inter-rater reliability are. Consider an experiment where two people are voting yes or no.

Intra-rater reliability is when the same type of experiment is completed by the same rater but in two or more different situations.

Agreement

Inter-rater reliability is when there are two different raters who are rating for the same experiment and they agree on the same vote.

Reliability

Understanding and evaluating Cohen’s Kappa

If there are N items that need to be classified into C mutually exclusive categories, the work of Cohen’s kappa is to measure the agreement between the two raters in order to classify N to C.

The value for this can be between 0 and 1 where 0 means there is no or random agreement between the raters and 1 indicates there is total agreement between them. But there can even be a negative value which indicates that there is absolutely no agreement between them.

To make things simple, let us derive the formula and make calculations to evaluate this metric.

6. Describe the model ensemble method. In machine leaming, what part does it play?

Ans:.

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model. To better understand this definition lets take a step back into ultimate goal of machine learning and model building. This is going to make more sense as I dive into specific examples and why Ensemble methods are used.

I will largely utilize Decision Trees to outline the definition and practicality of Ensemble Methods (however it is important to note that Ensemble Methods do not only pertain to Decision Trees).

A Decision Tree determines the predictive value based on series of questions and conditions. For instance, this simple Decision Tree determining on whether an individual should play outside or not. The tree takes several weather factors into account, and given each factor either makes a decision or asks another question. In this example, every time it is overcast, we will play outside. However, if it is raining, we must ask if it is windy or not? If windy, we will not play. But given no wind, tie those shoelaces tight because were going outside to play.Decision Trees can also solve quantitative problems as well with the same format. In the Tree to the left, we want to know wether or not to invest in a commercial real estate property.

7. What is a descriptive model's main purpose? Give examples of real-world Problems that descriptive models were used to solve.?

Ans:. Descriptive Models A descriptive model describes the domain it represents in a manner that can be interpreted by humans as well as computers. It can be used for many purposes, such as those described in Chapter 2, Section 2.2.2. It can include behavioral, structural, and other descriptions that establish logical relationships about the system, such as its whole-part relationship, the interconnection between its parts, and the allocation of its behavioral elements to structural elements. Descriptive models are generally not built in a manner that directly supports simulation, animation or execution, but they can be checked for consistency and adherence to the rules of the language, and the logical relationships can be reasoned about. The system model is a descriptive model that captures the requirements, structure, behavior, and parametric constraints associated with a system and its environment. The system model also captures inter-relationships between elements that represent its requirements, structure, behavior and parametric constraints. Because its modeling language supports various abstraction techniques, the system model also provides the ability to represent many other views of the system, such as a black-box view, white-box view, or a security view of the system. The system model can also be queried and analyzed for consistency, and serves as an integrating framework.

8. Describe how to evaluate a linear regression model.?

Ans:. Linear Regression Analysis consists of more than just fitting a linear line through a cloud of data points. It consists of 3 stages –

(1) analyzing the correlation and directionality of the data,

(2) estimating the model, i.e., fitting the line, and

(3) evaluating the validity and usefulness of the model.

First, a scatter plot should be used to analyze the data and check for directionality and correlation of data. The first scatter plot indicates a positive relationship between the two variables. The data is fit to run a regression analysis.

Linear Regression

Linear Regression

The second scatter plot seems to have an inverse U-shape this indicates that a regression line might not be the best way to explain the data, even if a correlation analysis establishes a positive link between the two variables.However, most often data contains quite a large amount of variability in these cases it is up for decision how to best proceed with the data.

Linear Regression

The first step enables the researcher to formulate the model, i.e. that variable X has a causal influence on variable Y and that their relationship is linear.

The second step of regression analysis is to fit the regression line. Mathematically least square estimation is used to minimize the unexplained residual. The basic idea behind this concept is illustrated in the following graph. In our example we want to model the relationship between age and job satisfaction. The research team has gathered several observations of self-reported job satisfaction and the age of the participant.

Linear Regression

When we fit a line through the scatter plot, the regression line represents the estimated job satisfaction for a given age. However the real observation might not fall exactly on the regression line. We try to explain the scatter plot with a linear equation of $y = b0 + b1x$. The distance between the regression line and the data point represents the unexplained variation, which is also called the residual ei.

The method of least squares is used to minimize the residual.